

# Supplementary information about the Turbo Medication Mapper (TMM), 2020-05-25

---

## SQL Query for Statin Drugs in our Clinical Data Warehouse (CDW)

---

*Also checks that they have been ordered for at least two unique patients. The schema and table name have been obscured.*

```
SELECT
  rm.PK_MEDICATION_ID,
  rm.FULL_NAME ,
  COUNT(DISTINCT pe.EMPI) AS empicount
FROM
  <MEDICATION TABLE> rm
LEFT JOIN ORDER_MED om ON
  rm.PK_MEDICATION_ID = om.FK_MEDICATION_ID
LEFT JOIN PATIENT_ENCOUNTER pe ON
  om.FK_PATIENT_ENCOUNTER_ID = pe.PK_PATIENT_ENCOUNTER_ID
WHERE
  (LOWER(FULL_NAME) LIKE '%statin%'
  OR LOWER(GENERIC_NAME) LIKE '%statin%'
  OR LOWER(FULL_NAME) LIKE '%lipitor%'
  OR LOWER(FULL_NAME) LIKE '%baycol%'
  OR LOWER(FULL_NAME) LIKE '%lescol%'
  OR LOWER(FULL_NAME) LIKE '%mevacor%'
  OR LOWER(FULL_NAME) LIKE '%livalo%'
  OR LOWER(FULL_NAME) LIKE '%pravachol%'
  OR LOWER(FULL_NAME) LIKE '%crestor%'
  OR LOWER(FULL_NAME) LIKE '%zypitamag%'
  OR LOWER(FULL_NAME) LIKE '%zocor%')
  AND LOWER(FULL_NAME) NOT LIKE '%leustatin%'
  AND LOWER(FULL_NAME) NOT LIKE '%cilastatin%'
  AND LOWER(FULL_NAME) NOT LIKE '%cholestatin%'
  AND LOWER(FULL_NAME) NOT LIKE '%sandostatin%'
  AND LOWER(FULL_NAME) NOT LIKE '%pentostatin%'
  AND LOWER(FULL_NAME) NOT LIKE '%nystatin%'
```

```

AND LOWER(GENERIC_NAME) NOT LIKE '%leustatin%'
AND LOWER(GENERIC_NAME) NOT LIKE '%cilastatin%'
AND LOWER(GENERIC_NAME) NOT LIKE '%cholestatin%'
AND LOWER(GENERIC_NAME) NOT LIKE '%sandostatin%'
AND LOWER(GENERIC_NAME) NOT LIKE '%pentostatin%'
AND LOWER(GENERIC_NAME) NOT LIKE '%nystatin%'
GROUP BY
    PK_MEDICATION_ID,
    FULL_NAME
HAVING
    COUNT(DISTINCT pe.EMPI) > 1

```

## Relation Between our CDW's **PHARMACY\_CLASS** and **PHARMACY\_SUBCLASS** Columns

We show **PHARMACY\_SUBCLASS**es for statins in our paper and mention **PHARMACY\_CLASS** and **THERAPEUTIC\_CLASS**. Here's a table that shows single-inheritance relationships from **PHARMACY\_CLASS** to **PHARMACY\_SUBCLASS**es with analgesics as an example.

PHARMACY_CLASS	PHARMACY_SUBCLASS	MEDICATIONs
Analgesics-nonnarcotic	Analgesic Combinations	1178
Analgesics-nonnarcotic	Analgesics Other	1610
Analgesics-nonnarcotic	Analgesics-Peptide Channel Blockers	11
Anti-rheumatic	Analgesics - Anti-inflammatory Combinations	68
Dermatological	Analgesics - Topical	382
Misc. genitourinary products	Urinary Analgesics	136
Otic	Otic Analgesics	27

## Confidential Information in Single-Patient **MEDICATION** Records

There are 1,464 **MEDICATION**s in our CDW whose **FULL\_NAME** contains the substring "william" and have been ordered for 0 or 1 patients, but 0 "william" orders that have been ordered for 2 or more patients. The same pattern holds for the substring "karen".

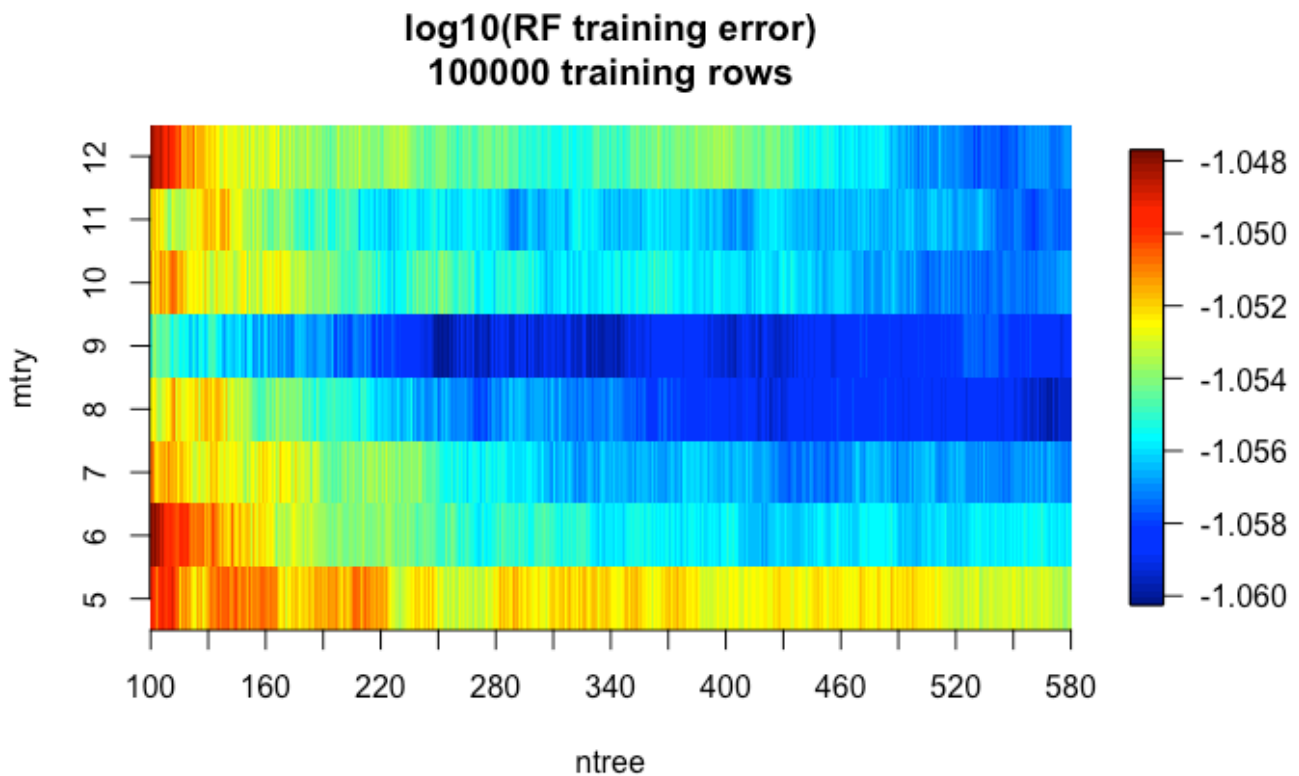
## Factor Importance for Random Forest (RF) Training

RxNav's `rank` and `score` were both retained, along with the `rxcul.count` relevance measure. The RxCUI `frequency` and the `counts` and `frequencies` of the highly granular RxNorm atoms were dropped due to low importance and/or > 0.95 correlation with `rxcul.count`. The largely orthogonal `qgram`, `jaccard`, `jw` and `cosine` string similarities are retrained, but `lcs` and `lv` were not.

feature	Mean Decrease Gini	Relative importance
score	14713.6	1.000
qgram	4837.0	0.329
rxcul.count	2668.4	0.181
rxcul.freq	2581.6	0.175
lcs	2370.5	0.161
jaccard	2325.7	0.158
jw	2297.0	0.156
q.char	2272.3	0.154
rank	2092.0	0.142
q.words	1815.1	0.123
sr.char	1707.1	0.116
cosine	1630.3	0.111
TTY.sr	1577.5	0.107
lv	1437.0	0.098
sr.words	1083.0	0.074
SAB.sr	736.8	0.050
rxcul.freq	247.8	0.017
rxcul.count	245.7	0.017

# Tuning of TMM RF

---



## Software Libraries Required for TMM's R Scripts

---

These may have further dependencies on other R libraries or even system software. Installing with R's `install.packages` function will generally resolve the R library dependency tree. `RJDBC` depends on `rJava`, which in turn requires Java to be installed on the system.

- RJDBC
- ROCR
- caret
- config
- dplyr
- fields
- ggplot2 (optional)
- httr
- jsonlite
- plyr
- randomForest
- rdflib
- readr
- solrium
- splitstackshape

- stringdist
- stringr
- tibble
- tidyr
- uuid